INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2009/M16063
February 2009,Lausanne,CH

Source: GIST (Gwangju Institute of Science and Technology)
Status: Proposal
Title: Experimental Results on Improved Temporal Consistency Enhancement
Author: Sang-Beom Lee, Cheon Lee, and Yo-Sung Ho

1. Introduction

This document reports experimental results on the improved temporal consistency enhancement for depth estimation. We improved the temporal consistency enhancement and implemented in Depth Estimation Reference Software (DERS) provided by Nagoya University. Then, we synthesized the intermediate views using View Synthesis Reference Software (VSRS) provided by Nagoya University. In order to evaluate the improved scheme, we compared the original view to the synthesized view.

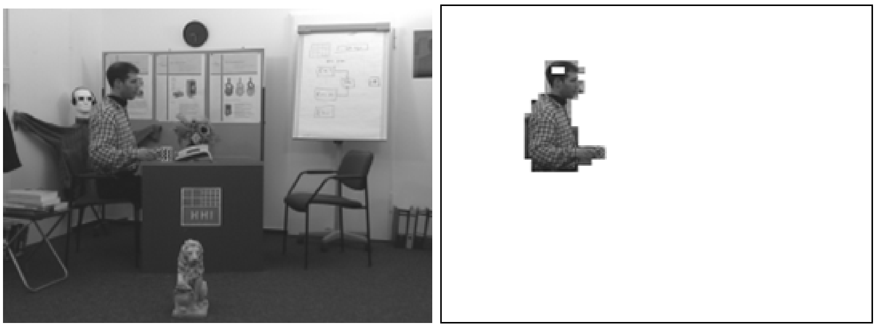
2. Improved Temporal Enhancement for Depth Estimation

In the previous work, we have noticed that the depth sequences had ghost artifacts near moving objects [1]. This problem was caused by the propagation of depth values at the previous frame and it degraded the visual quality of the synthesized views. Since the temporal weighting function refers to the previous depth value at exactly the same position, it is only worked well at the static background.
Therefore, we detect the moving object first. Since viewers feel the flickering artifacts at the background, we need to separate the moving object from the static background to apply the weighting function only to the background. We calculate mean absolute difference (MAD) for each block and distinguish by threshold whether the block is background or not. The temporal weighting function is defined by

$$C_{updated}(x,y,d) = C_{original}(x,y,d) + \lambda \, | d - D_{prev}(x,y) | \quad (1)$$

$$\lambda = \begin{cases} 1 & \text{if } MAD < Threshold \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $D_{prev}(x,y)$ represents the previous depth value. Figure 1 shows the result of the moving object detection.



(a) Original image (b) moving object detection

Fig. 1. Moving object detection

3. Experimental Results

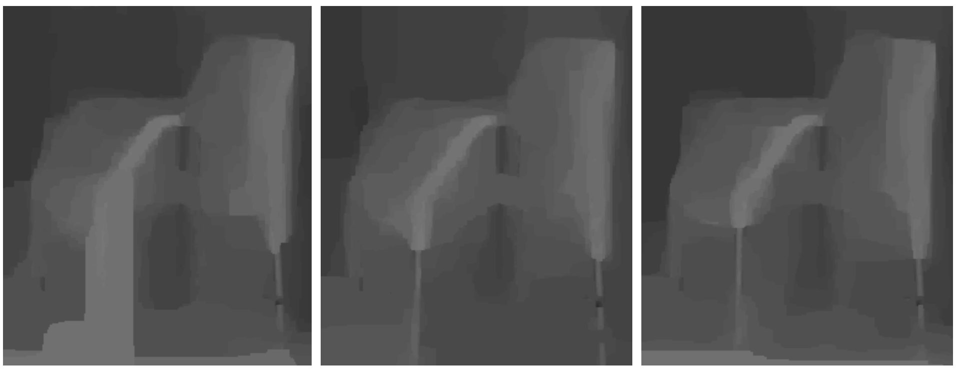
3.1. Depth Estimation Results

In order to evaluate our algorithm, we used ten test sequences: 'Book Arrival', 'Door Flower', 'Leaving Laptop', 'Alt Moabit', 'Champagne Tower', 'Pantomime', 'Dog', 'Newspaper', 'Lovebird1', and 'Lovebird2'. For depth estimation, we used Depth Estimation Reference Software (DERS). For moving object detection, we used the following parameter values as described in Table 1.

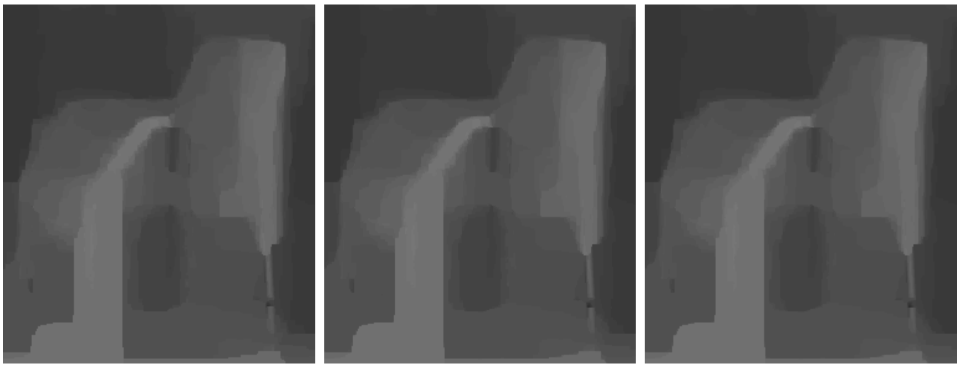
Table. 1. Threshold for moving object detection

Sequence	Threshold	Sequence	Threshold
Book Arrival	2.50	Pantomime	1.00
Door Flower	2.50	Dog	2.00
Leaving Laptop	2.50	Newspaper	1.50
Alt Moabit	2.50	Lovebird1	1.50
Champagne Tower	1.00	Lovebird2	1.50

Figure 2 shows the results of depth sequences for 'Book Arrival'. As shown in Fig. 2(a), we noticed that the depth sequence has inconsistent depth value at chair's legs, whereas the depth sequence in Fig.2(b) has consistent depth value.



(a) temporal consistency enhancement off



(b) temporal consistency enhancement on

Fig. 2. Results of depth sequences for three frames

3.2. View Synthesis Results

We used View Synthesis Reference Software (DERS) to synthesize the virtual view. Figure 3 shows the enlarged figures of virtual views for 'Book Arrival'. From the experimental results for 'Book Arrival', 'Door Flower' and 'Leaving Laptop', we noticed that the virtual views have inconsistent shapes as shown in Fig. 3(a), whereas the virtual views in Fig. 3(a) have consistent shape even if the shapes include errors at the chair's leg.

(a) temporal consistency enhancement off

(b) temporal consistency enhancement on

Fig. 3. Results of virtual views for three frames

Table 2 shows the results of PSNR comparison between the original view and the synthesized view. As shown in those results, PSNR for the results of temporal consistency enhancement was almost the same as that of the previous work. In other words, the flickering artifacts of synthesized views were reduced without any degradation of the objective quality.

Table. 2. Average PSNR

Sequence	View	Temporal consistency enhancement		Difference (ΔdB)
		Off (dB)	On (dB)	
Book Arrival	8	34.399	34.483	+0.084
Champagne Tower	39	28.766	28.719	-0.047
Dog	39	31.127	31.128	+0.001
Door Flower	8	36.160	36.203	+0.043
Leaving Laptop	8	36.066	35.984	-0.082
Lovebird1	6	30.987	31.000	+0.013
Lovebird2	6	34.182	34.196	+0.014
Alt Moabit	8	35.147	35.337	+0.190
Newspaper	4	24.373	24.370	-0.003
Pantomime	39	35.874	35.827	-0.047

4. Conclusion

We showed experimental results on the improved temporal consistency enhancement for depth estimation. We used the block-based moving object detection to separate the moving object from the static background and applied the temporal weighting function only at the background. We reduced the flickering artifacts of virtual views without any degradation of visual quality from the experimental results.

5. Acknowledgements

This work was supported in part by ITRC through RBRC at GIST (IITA-2009-C1090-0901-0017).

6. References

[1] ISO/IEC JTC1/SC29/WG11 “of Experiment on Temporal Enhancement for Depth Estimation,” M15852, October 2008.